

CAPÍTULO 20

BIG DATA E ONCOLOGIA DE PRECISÃO

Rodrigo Dienstmann

Diretor do programa de Medicina de Precisão – Grupo Oncoclínicas,
São Paulo

Introdução

O termo *big data* é definido como a área de conhecimento que estuda como obter, tratar e analisar informações a partir de conjuntos de dados distintos e grandes demais para serem analisados por sistemas tradicionais. O objetivo final do *big data* é agregar valor para os negócios ou, no caso em questão, para a Oncologia de Precisão. O conceito de *big data* é frequentemente vinculado aos 5 Vs: volume de dados, variedade de dados, velocidade de processamento, veracidade dos dados e valor dos dados obtidos (ou utilidade dos dados obtidos). Em Oncologia, o termo *big data* descreve amplamente a rápida aquisição e a geração de grandes quantidades de informações, geralmente de registros de câncer populacional, prontuário eletrônico, imagens médicas de alta resolução, biossensores com produção contínua de métricas fisiológicas ou estudos de sequenciamento genético em grande escala¹. Os limites da análise de tais dados apenas por humanos foram claramente excedidos, necessitando de uma maior dependência das máquinas.

O desafio de usar *big data* na pesquisa do câncer reside na colaboração interdisciplinar e no processamento de informações para unificar diversas fontes de dados e fornecer análises válidas para aproveitar informações significativas. Na Medicina, isso está começando a ter um impacto em três níveis: para os médicos, predominantemente por meio de interpretação rápida e precisa de imagens; para os sistemas de saúde, melhorando o fluxo de trabalho e o potencial de redução de erros médicos; e para os pacientes, permitindo-lhes processar seus próprios dados para promover a saúde. Ao mesmo tempo em que há mais dependência do que nunca de humanos para fornecer cuidados de saúde, algoritmos são necessários para ajudar na integração da inteligência humana e artificial para a Medicina de Precisão, que está apenas começando². Este capítulo fornece uma visão geral de como as abordagens de *big data* podem ser aplicadas na Oncologia de Precisão.

Decisões médicas baseadas em *big data*

Humanos gostam de regras. Grande parte das decisões médicas é baseada em regras construídas a partir de evidência científica. Com *big data* e tecnologia, as decisões passam a ser heurísticas e baseadas em modelos de *machine learning*³. Muitas vezes, o procedimento ou “algoritmo” que levou a uma resposta pode ser contraintuitivo e difícil de entender, mas, na prática clínica, a tecnologia para obter respostas a perguntas complexas não necessita ser perfeita, apenas suficientemente boa em condições reais. O *big data* em Medicina é a maneira moderna de resolver problemas: ciência além da arte, precisão além da intuição.

Todas as tecnologias disruptivas, como os sistemas de suporte à decisão baseados em inteligência artificial, passam por um momento de “*hype*” ou expectativas infladas em que a maioria dos humanos fica inicialmente frustrada com sua baixa eficiência na prática. Mas o crescimento tecnológico exponencial

que ocorre quando máquinas são capazes de aprender com grandes quantidades de dados gera uma revolução na produtividade e consequentemente maior aceitação. Dessa forma, o impacto tecnológico de *big data* e de modelos de inteligência artificial é frequentemente superestimado por humanos na etapa inicial de desenvolvimento e de aplicação prática, mas nitidamente subestimado em longo prazo, na fase final de implementação.

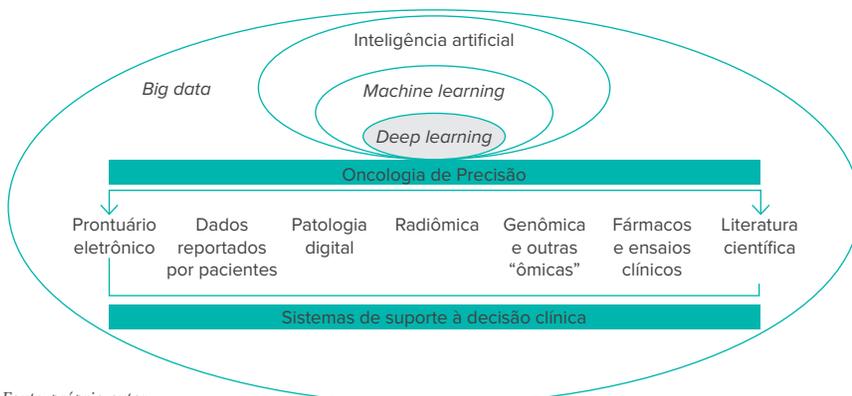
Etapas para gerar valor através de *big data*

Para que os dados possam gerar valor e uma aplicação clínica real, várias etapas devem ser seguidas⁴. No campo da Oncologia de Precisão isso implica em integração de dados clínicos, patológicas e multiômicos (**volume** e **variedade**), por exemplo, no desenho de um algoritmo com potencial diagnóstico, prognóstico ou preditivo. É muito importante garantir que dados utilizados para treinar os modelos de *machine learning* sejam completos, estejam em conformidade com padrões de formato para sua geração e que sejam plausíveis de acordo com a avaliação de especialistas na área (**veracidade**). Ou seja, os algoritmos devem ser treinados em dados curados por humanos, que servem como padrão-ouro na definição do desfecho a ser investigado. A coleta, a análise e o resultado final devem ser obtidos de maneira rápida e sem gargalos (**velocidade**). Em seguida, os modelos devem passar por validação externa, com dados até então não utilizados no desenvolvimento original do algoritmo. Esse processo é crítico para a validação clínica e para a avaliação da performance do modelo no mundo real (**valor**), controlando sobreajustes do modelo aos dados originais. Finalmente, é fundamental que os algoritmos gerem um resultado de fácil entendimento (resposta simples a uma pergunta complexa) e que a confiança (ou margem de erro) na estimativa de uma classificação também seja apresentada, para facilitar uma aplicação clínica e a tomada de decisão por parte de humanos. Contudo, a validação do desempenho de um algoritmo em termos de sua precisão não é equivalente a demonstrar eficácia clínica. Ou seja, um algoritmo com AUC (*area under the curve*) de 0,99 pode não ser válido se não provar melhoria nos resultados clínicos frente ao diagnóstico padrão². Em resumo, para desenvolver um aplicativo baseado em *big data* e inteligência artificial, é necessário: (1) formular a pergunta correta no contexto apropriado (relevância clínica e acesso a dados de qualidade); (2) avaliar rigorosamente o modelo (validação externa, definição de sucesso e facilidade de implementação clínica); (3) preocupar-se com a transição ao mercado (com uma licença para uso de *software* ou aplicativo, por exemplo); e (4) trabalhar em equipes interdisciplinares com especialistas em Medicina, Bioinformática, Ciências de Dados, além dos usuários, administradores e agências regulatórias⁴.

Existem inúmeros modelos preditivos que podem ser aplicados em *big data*, desde regressão linear (que costuma ser de menor acurácia, mas maior

interpretabilidade) até *deep neural networks* (de maior acurácia e menor interpretabilidade). O *deep learning* é o novo paradigma para a realização do *machine learning* e ambos são pilares que sustentam a inteligência artificial. Uma das principais diferenças é que o *deep learning* é intuitivo e funciona como uma mente própria através de sobreposição de camadas não lineares de processamento de dados, enquanto o *machine learning* exige uma intervenção manual na seleção dos recursos a serem processados². Com a tecnologia atual, os modelos são capazes de auxiliar na resolução de apenas uma tarefa, o que também é chamado de *shallow artificial intelligence*. Um exemplo seria um algoritmo para classificar mutações no genoma de um paciente em patogenias de significado incerto ou benignas. Isso é infinitamente mais simples que modelos que estão desenhados para dar respostas a múltiplas perguntas simultaneamente e com alta acurácia, chamados de *deep artificial intelligence*. Ou seja, a inteligência artificial atual, que será utilizada em Medicina de Precisão nos próximos anos, não objetiva substituir a intuição e o raciocínio humanos necessários para responder a múltiplas perguntas e tomar decisões na vida real. Contudo, existe absoluta necessidade de utilizar inteligência artificial no campo da Oncologia de Precisão. Humanos são capazes de avaliar em média sete elementos (dados) em paralelo dentro de um processo cognitivo⁵. Se o médico necessita integrar informações relacionadas às características clínicas do paciente, sua história de tratamentos prévios, múltiplos exames laboratoriais e de imagem consecutivos, marcadores patológicos e genômicos do tumor, interpretar a genética do indivíduo, identificar as alterações patogênicas e acionáveis, priorizar tratamentos dirigidos com base em descobertas científicas recentes e reconhecer as alternativas de terapias em ensaios clínicos, para então finalmente tomar uma decisão assertiva e individualizada, nesse contexto um algoritmo capaz de guiar a decisão clínica é de grande valor. A figura 1 ilustra esse conceito.

Figura 1. Esquema resumindo intersecção de *big data* e inteligência artificial aplicados à Oncologia de Precisão, com o objetivo final de gerar impacto na tomada de decisões clínica



Fonte: próprio autor.

A Medicina é tão ampla e complexa que é difícil – ou praticamente impossível – capturar a informação relevante em regras. Nesse contexto, se o algoritmo tem valor real em dar respostas a perguntas relevantes, mesmo que seja de difícil interpretação (*black box*), ainda assim pode ter aplicabilidade clínica. Um exemplo seria a utilização de algoritmos de inteligência artificial para diferenciar um nódulo benigno de lesão maligna em exames de imagem pulmonar de alta resolução. Alguns detalhes podem passar despercebidos ao olho humano e máquinas são capazes de reconhecer padrões únicos que aumentam a acurácia diagnóstica. Até o momento, em Oncologia, os algoritmos de inteligência artificial aprovados para uso clínico por agências regulatórias são aplicados a dados de imagem, tanto de Radiologia (para melhor definir o risco de lesão maligna em uma ressonância magnética de mama) como em Patologia (para acelerar a revisão de lâminas de tecido prostático ou mamário e identificar células malignas de maneira automática)⁶.

Big data em imagens de Radiologia e Patologia

Imagens médicas quantitativas fornecem dados de fenótipo tumoral, incluindo forma, tamanho, volume e textura do tumor. Esses recursos podem ser correlacionados com dados clínicos e usados para o suporte de decisão clínica em conjunto com outras informações fornecidas por relatórios clínicos, como dados “ômicos” e testes de laboratório.

Os dados de imagens médicas de Radiologia são comumente gerados utilizando técnicas como tomografia computadorizada (TC), imagem por ressonância magnética (RNM) ou tomografia por emissão de pósitrons (PET). Incorporar as informações da anatomia e da atividade metabólica ajuda a distinguir melhor nódulos ou massas benignas das malignas encontradas nas imagens.

Veja esse conteúdo na íntegra
acessando a OC Academia:
<http://ocacademia.com>

